**Project 4 Proposal**

Introduction:

* Key topic will be on mass shooting in United States.
* To use multiple sources and data science to verify variables that could correlate with mass shooting

Problem Statement:

* Do the states that provides more education opportunities (more schools) to the people have less mass shooting case?
* Does mass shooting have correlation with suicidal rate?
* Is there any time trend with the shooter’s age?

EDA:

* Determine states that are more prone to mass shooting
* Time trend with number of victims
* Determine the year with highest number of mass shooting & number of victims
* Ratio of target (Specified target vs Random strangers)
* General overview of all available data

Datasets available:

1. US Mass Shooting

* **Columns:** S#; Title; Location; Date; Incident Area; Open/Close Location; Target; Cause; Summary; Fatalities; Injured; Total victims; Policeman Killed; Age; Employed (Y/N); Employed at; Mental Health Issues; Race; Gender; Latitude; Longitude
* **Number of rows:** 323

1. Number of schools in United States

<https://nces.ed.gov/pubs2007/overview04/tables/table_2.asp>

1. Suicidal rate by country

* **Columns**: Entity Code; Year; suicide rate (age-adjusted suicides per 100,000 people)
* **Number of rows**: 2843